

Madeleine Thompson
DSCI 551

HW5 – Hadoop MapReduce & Spark

2. (a) `spark-submit dataframe-a.py country.json`

Name
Aruba
Anguilla
Netherlands Antilles
Antigua and Barbuda
Bahamas
Belize
Bermuda
Barbados
Canada
Costa Rica
Cuba
Cayman Islands
Dominica
Dominican Republic
Guadeloupe
Grenada
Greenland
Guatemala
Honduras
Haiti

only showing top 20 rows

RDD: `spark-submit rdd-a.py country.json`

```
['Antigua and Barbuda', 'Costa Rica', 'Cuba', 'Cayman Islands', 'Dominica', 'Honduras', 'Haiti', 'Mexico', 'Puerto Rico', 'Trinidad and Tobago', 'Bahamas', 'Belize', 'Bermuda', 'Dominican Republic', 'Jamaica', 'Montserrat', 'Martinique', 'United States', 'Saint Vincent and the Grenadines', 'Virgin Islands, British', 'Virgin Islands, U.S.', 'Barbados', 'Canada', 'Saint Kitts and Nevis', 'Nicaragua', 'El Salvador', 'Saint Pierre and Miquelon', 'Aruba', 'Anguilla', 'Netherlands Antilles', 'Guadeloupe', 'Grenada', 'Greenland', 'Guatemala', 'Saint Lucia', 'Panama', 'Turks and Caicos Islands']
```

(b) spark-submit dataframe-b.py country.json city.json

CountryName	Name
Afghanistan	Kabul
Netherlands	Amsterdam
Netherlands Antilles	Willemstad
Albania	Tirana
Algeria	Alger
American Samoa	Fagatogo
Andorra	Andorra la Vella
Angola	Luanda
Anguilla	The Valley
Antigua and Barbuda	Saint John's
United Arab Emirates	Abu Dhabi
Argentina	Buenos Aires
Armenia	Yerevan
Aruba	Oranjestad
Australia	Canberra
Azerbaijan	Baku
Bahamas	Nassau
Bahrain	al-Manama
Bangladesh	Dhaka
Barbados	Bridgetown

only showing top 20 rows

RDD: spark-submit rdd-b.py country.json city.json

```
[('Azerbaijan', 'Baku'), ('Burundi', 'Bujumbura'), ('Costa Rica', 'San JosÃ©'), ('Guyana', 'Georgeto  
wn'), ('Poland', 'Warszawa'), ('French Polynesia', 'Papeete'), ('Trinidad and Tobago', 'Port-of-Spai  
n'), ('Tuvalu', 'Funafuti'), ('Belarus', 'Minsk'), ('Egypt', 'Cairo'), ('Italy', 'Roma'), ('Mongolia  
, 'Ulan Bator'), ('Malaysia', 'Kuala Lumpur'), ('Singapore', 'Singapore'), ('Bhutan', 'Thimphu'), ('  
'Switzerland', 'Bern'), ('United Kingdom', 'London'), ('Kazakhstan', 'Astana'), ('Cambodia', 'Phnom P  
enh'), ('Kiribati', 'Bairiki'), ('Saint Kitts and Nevis', 'Basseterre'), ('Laos', 'Vientiane'), ('Wa  
llis and Futuna', 'Mata-Utu'), ('Yugoslavia', 'Beograd'), ('Angola', 'Luanda'), ('Congo', 'Brazzavil  
le'), ('Gambia', 'Banjul'), ('Liberia', 'Monrovia'), ('Nauru', 'Yaren'), ('Pitcairn', 'Adamstown'),  
( 'Sweden', 'Stockholm'), ('Thailand', 'Bangkok'), ('Cayman Islands', 'George Town'), ('Djibouti', 'D  
jibouti'), ('Guam', 'Agaña'), ('Hong Kong', 'Victoria'), ('Croatia', 'Zagreb'), ('Haiti', 'Port-au-  
Prince'), ('RÃ©union', 'Saint-Denis'), ('Uganda', 'Kampala'), ('Bosnia and Herzegovina', 'Sarajevo')  
, ('Belize', 'Belmopan'), ('Iceland', 'ReykjavÃ­xadk'), ('Northern Mariana Islands', 'Garapan'), ('M  
ontserrat', 'Plymouth'), ('Sudan', 'Khartum'), ('Virgin Islands, British', 'Road Town'), ('Central A  
frican Republic', 'Bangui'), ('Micronesia, Federated States of', 'Palikir'), ('Georgia', 'Tbilisi'),  
( 'Kenya', 'Nairobi'), ('Solomon Islands', 'Honiara'), ('Slovakia', 'Bratislava'), ('Vanuatu', 'Port  
-Vila'), ('Samoa', 'Apia'), ('Aruba', 'Oranjestad'), ('Afghanistan', 'Kabul'), ('Netherlands Antille  
s', 'Willemstad'), ('United Arab Emirates', 'Abu Dhabi'), ('Colombia', 'SantaFÃ© de BogotÃ¡'), ('Gre  
ece', 'Athenai'), ('Libyan Arab Jamahiriya', 'Tripoli'), ('Saint Lucia', 'Castries'), ('Sri Lanka',  
'Colombo'), ('Nepal', 'Kathmandu'), ('Palau', 'Koror'), ('Chad', 'NÃ'DjamÃ©na'), ('Dominica', 'Rosea  
u'), ('Moldova', 'Chisinau'), ('Mali', 'Bamako'), ('Portugal', 'Lisboa'), ('Palestine', 'Gaza'), ('E  
ast Timor', 'Dili'), ('Tanzania', 'Dodoma'), ('Bolivia', 'La Paz'), ('Ecuador', 'Quito'), ('Israel',  
'Jerusalem'), ('Jamaica', 'Kingston'), ('Jordan', 'Amman'), ('Mozambique', 'Maputo'), ('Romania', 'B  
ucuresti'), ('Svalbard and Jan Mayen', 'Longyearbyen'), ('Ukraine', 'Kyiv'), ('Holy See (Vatican Ci  
ty State)', 'CittÃx0 del Vaticano'), ('Saint Vincent and the Grenadines', 'Kingstown'), ('Brunei',  
'Bandar Seri Begawan'), ('Mayotte', 'Mamoutzou'), ('Niger', 'Niamey'), ('Nigeria', 'Abuja'), ('Viet  
nam', 'Hanoi'), ('Zambia', 'Lusaka'), ('Albania', 'Tirana'), ('Chile', 'Santiago de Chile'), ('Congo  
, The Democratic Republic of the', 'Kinshasa'), ('Guatemala', 'Ciudad de Guatemala'), ('Latvia', 'Ri  
ga'), ('Panama', 'Ciudad de PanamÃ¡'), ('Peru', 'Lima'), ('Syria', 'Damascus'), ('Austria', 'Wien'),  
( 'Belgium', 'Bruxelles [Brussel]'), ('Benin', 'Porto-Novo'), ('Czech Republic', 'Praha'), ('Denmark  
, 'KÃbenhavn'), ('Hungary', 'Budapest'), ('Indonesia', 'Jakarta'), ('Mexico', 'Ciudad de MÃxico'),  
( 'Marshall Islands', 'Dalap-Uliga-Darrit'), ('Turkmenistan', 'Ashgabat'), ('Bulgaria', 'Sofija'),  
( 'Dominican Republic', 'Santo Domingo de GuzmÃ¡n'), ('Algeria', 'Alger'), ('Venezuela', 'Caracas'),  
( 'Virgin Islands, U.S.', 'Charlotte Amalie'), ('Brazil', 'BrasÃxadia'), ('Falkland Islands', 'Stan  
ley'), ('Gibraltar', 'Gibraltar'), ('South Korea', 'Seoul'), ('San Marino', 'San Marino'), ('Saint P  
ierre and Miquelon', 'Saint-Pierre'), ('Suriname', 'Paramaribo'), ('China', 'Peking'), ('Cape Verde',  
'Praia'), ('New Zealand', 'Wellington'), ('Germany', 'Berlin'), ('Macedonia', 'Skopje'), ('Malta',  
'Valletta'), ('Papua New Guinea', 'Port Moresby'), ('Bahamas', 'Nassau'), ('Eritrea', 'Asmara'), ('  
Ethiopia', 'Addis Abeba'), ('Finland', 'Helsinki [Helsingfors]'), ('Iran', 'Teheran'), ('Japan', 'To  
kyo'), ('Martinique', 'Fort-de-France'), ('Russian Federation', 'Moscow'), ('Uruguay', 'Montevideo'),  
( 'Botswana', 'Gaborone'), ('Fiji Islands', 'Suva'), ('Sao Tome and Principe', 'SÃo TomÃ©'), ('Slo  
venia', 'Ljubljana'), ('Yemen', 'Sanaa'), ('South Africa', 'Pretoria'), ('Zimbabwe', 'Harare'), ('Ca  
meroon', 'YaoundÃ©'), ('Equatorial Guinea', 'Malabo'), ('Grenada', 'Saint George's'), ('Luxembourg',  
'Luxembourg [Luxemburg/LÃtzebuerg]'), ('Swaziland', 'Mbabane'), ('Togo', 'LomÃ©'), ('Cuba', 'La H  
abana'), ('Honduras', 'Tegucigalpa'), ('India', 'New Delhi'), ('Paraguay', 'AsunciÃ³n'), ('Qatar', 'Doha'),  
( 'Tunisia', 'Tunis'), ('Burkina Faso', 'Ouagadougou'), ('Bahrain', 'al-Manama'), ('Western S  
ahara', 'EL-AaiÃn'), ('Spain', 'Madrid'), ('Iraq', 'Baghdad'), ('Mauritania', 'Nouakchott'), ('Saud  
i Arabia', 'Riyadh'), ('United States', 'Washington'), ('Cocos (Keeling) Islands', 'West Island'), ('
```

(c) spark-submit dataframe-c.py country.json

```
+-----+  
|  Continent|  
+-----+  
|      Europe|  
|      Africa|  
|North America|  
|    Antarctica|  
|South America|  
|      Oceania|  
|        Asia|  
+-----+
```

(d) spark-submit dataframe-d.py countrylanguage.json

```
+-----+  
|   Language|  
+-----+  
|      Chinese|  
|       Dutch|  
|     English|  
|Eskimo Languages|  
|      French|  
|      German|  
|      Italian|  
|      Polish|  
|Portuguese|  
|      Punjabi|  
|      Spanish|  
|    Ukrainian|  
+-----+
```

(e) spark-submit dataframe-e.py country.json

```
+-----+-----+  
|Continent|    avg_le|  
+-----+-----+  
|    Africa|51.6655172413793|  
+-----+-----+
```